

RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.

Application Serial Number: 10/787,267A
Source: FWP
Date Processed by STIC: 10/22/04

ENTERED



IFWO

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/787,267A

DATE: 10/22/2004

TIME: 12:50:01

Input Set : A:\GC687-3-D1-seqlist-corr.txt

Output Set: N:\CRF4\10222004\J787267A.raw

```

4 <110> APPLICANT: Dartois, Veronique A.
5     Hoch, James A.
6     Valle, Fernando
7     Kumar, Manoj
9 <120> TITLE OF INVENTION: 2,5-DKG Permeases
12 <130> FILE REFERENCE: GC687-3-D1
14 <140> CURRENT APPLICATION NUMBER: US 10/787,267A
15 <141> CURRENT FILING DATE: 2004-02-25
17 <150> PRIOR APPLICATION NUMBER: US 09/922,501
18 <151> PRIOR FILING DATE: 2001-08-03
20 <150> PRIOR APPLICATION NUMBER: US 60/325,774
21 <151> PRIOR FILING DATE: 2000-08-04
23 <150> PRIOR APPLICATION NUMBER: US 60/421,141
24 <151> PRIOR FILING DATE: 2000-09-29
26 <160> NUMBER OF SEQ ID NOS: 22
28 <170> SOFTWARE: FastSEQ for Windows Version 4.0
30 <210> SEQ ID NO: 1
31 <211> LENGTH: 1500
32 <212> TYPE: DNA
33 <213> ORGANISM: Unknown
35 <220> FEATURE:
36 <223> OTHER INFORMATION: environmental source
39 <220> FEATURE:
40 <221> NAME/KEY: CDS
41 <222> LOCATION: (94)...(1374)
43 <400> SEQUENCE: 1
44 ggcgaatagc cgggccggcg tcataataac ggccttctct gtaccctaca tacggcgggcg      60
45 gcgtcatgaa cctcaacttt agtaggcaag cct atg aac agc tct acc aat gca      114
46                                     Met Asn Ser Ser Thr Asn Ala
47                                     1           5
49 acg aaa cgc tgg tgg tac atc atg cct atc gtg ttt atc acg tat agc      162
50 Thr Lys Arg Trp Trp Tyr Ile Met Pro Ile Val Phe Ile Thr Tyr Ser
51     10           15           20
53 ctg gcg tat ctc gac cgc gca aac ttc agc ttt gct tcg gca gcg ggc      210
54 Leu Ala Tyr Leu Asp Arg Ala Asn Phe Ser Phe Ala Ser Ala Ala Gly
55     25           30           35
57 att acg gaa gat tta ggc att acc aaa ggc atc tcg tcg ctt ctt ggc      258
58 Ile Thr Glu Asp Leu Gly Ile Thr Lys Gly Ile Ser Ser Leu Leu Gly
59 40           45           50           55
61 gca ctt ttc ttc ctc ggc tat ttc ttc ttc cag atc ccg ggg gcg att      306
62 Ala Leu Phe Phe Leu Gly Tyr Phe Phe Phe Gln Ile Pro Gly Ala Ile
63     60           65           70
65 tac gcg gaa cgc cgt agc gta cgg aag ctg att ttc atc tgt ctg atc      354

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/787,267A

DATE: 10/22/2004

TIME: 12:50:01

Input Set : A:\GC687-3-D1-seqlist-corr.txt

Output Set: N:\CRF4\10222004\J787267A.raw

66	Tyr	Ala	Glu	Arg	Arg	Ser	Val	Arg	Lys	Leu	Ile	Phe	Ile	Cys	Leu	Ile	
67				75					80					85			
69	ctg	tgg	ggc	gcc	tgc	gcc	tcg	ctt	gac	cgg	gat	ggt	gca	caa	tat	tcc	402
70	Leu	Trp	Gly	Ala	Cys	Ala	Ser	Leu	Asp	Arg	Asp	Gly	Ala	Gln	Tyr	Ser	
71			90					95					100				
73	agc	gct	ggc	tgg	cga	tcc	gtt	tta	ttc	tcg	gct	gtc	gtg	gaa	gcg	gcg	450
74	Ser	Ala	Gly	Trp	Arg	Ser	Val	Leu	Phe	Ser	Ala	Val	Val	Glu	Ala	Ala	
75		105					110					115					
77	gtc	atg	ccg	gcg	atg	ctg	att	tac	atc	agt	aac	tgg	ttt	acc	aaa	tca	498
78	Val	Met	Pro	Ala	Met	Leu	Ile	Tyr	Ile	Ser	Asn	Trp	Phe	Thr	Lys	Ser	
79	120					125					130					135	
81	gaa	cgt	tca	cgc	gcc	aac	acc	ttc	tta	atc	ctc	ggc	aac	ccg	gtc	acg	546
82	Glu	Arg	Ser	Arg	Ala	Asn	Thr	Phe	Leu	Ile	Leu	Gly	Asn	Pro	Val	Thr	
83				140					145				150				
85	gta	ctg	tgg	atg	tcg	gtg	gtc	tcc	ggc	tac	ctg	att	cag	tcc	ttc	ggc	594
86	Val	Leu	Trp	Met	Ser	Val	Val	Ser	Gly	Tyr	Leu	Ile	Gln	Ser	Phe	Gly	
87			155					160				165					
89	tgg	cgt	gaa	atg	ttt	att	att	gaa	ggc	gtt	ccg	gcc	gtc	ctc	tgg	gcc	642
90	Trp	Arg	Glu	Met	Phe	Ile	Ile	Glu	Gly	Val	Pro	Ala	Val	Leu	Trp	Ala	
91			170				175					180					
93	ttc	tgc	tgg	tgg	gtg	ctg	gtc	aaa	gtt	aaa	ccg	tcg	cag	gtg	aac	tgg	690
94	Phe	Cys	Trp	Trp	Val	Leu	Val	Lys	Val	Lys	Pro	Ser	Gln	Val	Asn	Trp	
95		185				190					195						
97	ttg	tcg	gaa	aac	gag	aaa	gcc	gcg	ctg	cag	gcg	cag	ctg	gag	agc	gag	738
100	Leu	Ser	Glu	Asn	Glu	Lys	Ala	Ala	Leu	Gln	Ala	Gln	Leu	Glu	Ser	Glu	
101	200				205				210				215				
103	cag	cag	ggt	att	aaa	gcc	gtg	cgt	aac	tac	ggc	gaa	gcc	ttc	cgc	tca	786
104	Gln	Gln	Gly	Ile	Lys	Ala	Val	Arg	Asn	Tyr	Gly	Glu	Ala	Phe	Arg	Ser	
105				220					225				230				
107	cgc	aac	gtc	att	cta	ctg	tgc	atg	cag	tat	ttt	gcc	tgg	agt	atc	ggc	834
108	Arg	Asn	Val	Ile	Leu	Leu	Cys	Met	Gln	Tyr	Phe	Ala	Trp	Ser	Ile	Gly	
109			235					240				245					
111	gtg	tac	ggt	ttt	gtg	ctg	tgg	ttg	ccg	tca	att	att	cgc	agc	ggc	ggc	882
112	Val	Tyr	Gly	Phe	Val	Leu	Trp	Leu	Pro	Ser	Ile	Ile	Arg	Ser	Gly	Gly	
113			250				255					260					
115	gtc	aat	atg	ggg	atg	gtg	gaa	gtc	ggc	tgg	ctc	tct	tcg	gtg	cct	tat	930
116	Val	Asn	Met	Gly	Met	Val	Glu	Val	Gly	Trp	Leu	Ser	Ser	Val	Pro	Tyr	
117		265				270					275						
119	ctg	gcc	gcg	act	att	gcg	atg	atc	gtc	gtc	tcc	tgg	gct	tcc	gat	aaa	978
120	Leu	Ala	Ala	Thr	Ile	Ala	Met	Ile	Val	Val	Ser	Trp	Ala	Ser	Asp	Lys	
121	280				285					290							
123	atg	cag	aac	cgt	aaa	ctg	ttc	gtc	tgg	ccg	ctg	ctg	ctg	att	ggc	gga	1026
124	Met	Gln	Asn	Arg	Lys	Leu	Phe	Val	Trp	Pro	Leu	Leu	Leu	Ile	Gly	Gly	
125			300					305				310					
127	ctg	gct	ttt	att	ggc	tca	tgg	gcc	gtc	ggc	gct	aac	cat	ttc	tgg	gcc	1074
128	Leu	Ala	Phe	Ile	Gly	Ser	Trp	Ala	Val	Gly	Ala	Asn	His	Phe	Trp	Ala	
129			315				320					325					
131	tct	tat	acc	ctg	ctg	gtg	att	gcc	aat	gcg	gca	atg	tac	gcc	cct	tac	1122
132	Ser	Tyr	Thr	Leu	Leu	Val	Ile	Ala	Asn	Ala	Ala	Met	Tyr	Ala	Pro	Tyr	

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/787,267A

DATE: 10/22/2004

TIME: 12:50:01

Input Set : A:\GC687-3-D1-seqlist-corr.txt

Output Set: N:\CRF4\10222004\J787267A.raw

```

133          330          335          340
135 ggt ccg ttt ttc gcc atc att ccg gaa atg ctg ccg cgt aac gtc gcc      1170
136 Gly Pro Phe Phe Ala Ile Ile Pro Glu Met Leu Pro Arg Asn Val Ala
137          345          350          355
139 ggt ggc gca atg gcg ctc atc aac agc atg ggg gcc tta ggt tca ttc      1218
140 Gly Gly Ala Met Ala Leu Ile Asn Ser Met Gly Ala Leu Gly Ser Phe
141 360          365          370          375
143 ttt ggt tcg tgg ttc gtg ggc tac ctg aac ggc acc acc ggc agt cca      1266
144 Phe Gly Ser Trp Phe Val Gly Tyr Leu Asn Gly Thr Thr Gly Ser Pro
145          380          385          390
147 tca gcc tca tac att ttc atg gga gtg gcg ctt ttc gcc tcg gta tgg      1314
148 Ser Ala Ser Tyr Ile Phe Met Gly Val Ala Leu Phe Ala Ser Val Trp
149          395          400          405
151 ctt act tta att gtt aag cct gct aac aat caa aag ctc ccc atc ggc      1362
152 Leu Thr Leu Ile Val Lys Pro Ala Asn Asn Gln Lys Leu Pro Ile Gly
153          410          415          420
155 gct cgt cac gcc tgacctttac tacttacgga gatcacgcct tgggtacgtt      1414
156 Ala Arg His Ala
157          425
159 gcaggacaaa ccgataggca ccgcaaaggc tggggccatc gagcagcgcg taaacagtca      1474
160 gctggttgct gtcgctgtgc ggcgtc      1500
162 <210> SEQ ID NO: 2
163 <211> LENGTH: 427
164 <212> TYPE: PRT
165 <213> ORGANISM: Unknown
167 <220> FEATURE:
168 <223> OTHER INFORMATION: environmental source
170 <400> SEQUENCE: 2
171 Met Asn Ser Ser Thr Asn Ala Thr Lys Arg Trp Trp Tyr Ile Met Pro
172 1          5          10          15
173 Ile Val Phe Ile Thr Tyr Ser Leu Ala Tyr Leu Asp Arg Ala Asn Phe
174          20          25          30
175 Ser Phe Ala Ser Ala Ala Gly Ile Thr Glu Asp Leu Gly Ile Thr Lys
176          35          40          45
177 Gly Ile Ser Ser Leu Leu Gly Ala Leu Phe Phe Leu Gly Tyr Phe Phe
178          50          55          60
179 Phe Gln Ile Pro Gly Ala Ile Tyr Ala Glu Arg Arg Ser Val Arg Lys
180 65          70          75          80
181 Leu Ile Phe Ile Cys Leu Ile Leu Trp Gly Ala Cys Ala Ser Leu Asp
182          85          90          95
183 Arg Asp Gly Ala Gln Tyr Ser Ser Ala Gly Trp Arg Ser Val Leu Phe
184          100          105          110
185 Ser Ala Val Val Glu Ala Ala Val Met Pro Ala Met Leu Ile Tyr Ile
186          115          120          125
187 Ser Asn Trp Phe Thr Lys Ser Glu Arg Ser Arg Ala Asn Thr Phe Leu
188          130          135          140
189 Ile Leu Gly Asn Pro Val Thr Val Leu Trp Met Ser Val Val Ser Gly
190 145          150          155          160
191 Tyr Leu Ile Gln Ser Phe Gly Trp Arg Glu Met Phe Ile Ile Glu Gly

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/787,267A

DATE: 10/22/2004

TIME: 12:50:01

Input Set : A:\GC687-3-D1-seqlist-corr.txt

Output Set: N:\CRF4\10222004\J787267A.raw

```

192                               165                               170                               175
193 Val Pro Ala Val Leu Trp Ala Phe Cys Trp Trp Val Leu Val Lys Val
194                               180                               185                               190
195 Lys Pro Ser Gln Val Asn Trp Leu Ser Glu Asn Glu Lys Ala Ala Leu
196                               195                               200                               205
197 Gln Ala Gln Leu Glu Ser Glu Gln Gln Gly Ile Lys Ala Val Arg Asn
198                               210                               215                               220
199 Tyr Gly Glu Ala Phe Arg Ser Arg Asn Val Ile Leu Leu Cys Met Gln
200 225                               230                               235                               240
201 Tyr Phe Ala Trp Ser Ile Gly Val Tyr Gly Phe Val Leu Trp Leu Pro
202                               245                               250                               255
203 Ser Ile Ile Arg Ser Gly Gly Val Asn Met Gly Met Val Glu Val Gly
204                               260                               265                               270
205 Trp Leu Ser Ser Val Pro Tyr Leu Ala Ala Thr Ile Ala Met Ile Val
206                               275                               280                               285
207 Val Ser Trp Ala Ser Asp Lys Met Gln Asn Arg Lys Leu Phe Val Trp
208                               290                               295                               300
209 Pro Leu Leu Leu Ile Gly Gly Leu Ala Phe Ile Gly Ser Trp Ala Val
210 305                               310                               315                               320
211 Gly Ala Asn His Phe Trp Ala Ser Tyr Thr Leu Leu Val Ile Ala Asn
212                               325                               330                               335
213 Ala Ala Met Tyr Ala Pro Tyr Gly Pro Phe Phe Ala Ile Ile Pro Glu
214                               340                               345                               350
215 Met Leu Pro Arg Asn Val Ala Gly Gly Ala Met Ala Leu Ile Asn Ser
216                               355                               360                               365
217 Met Gly Ala Leu Gly Ser Phe Phe Gly Ser Trp Phe Val Gly Tyr Leu
218                               370                               375                               380
219 Asn Gly Thr Thr Gly Ser Pro Ser Ala Ser Tyr Ile Phe Met Gly Val
220 385                               390                               395                               400
221 Ala Leu Phe Ala Ser Val Trp Leu Thr Leu Ile Val Lys Pro Ala Asn
222                               405                               410                               415
223 Asn Gln Lys Leu Pro Ile Gly Ala Arg His Ala
224                               420                               425
226 <210> SEQ ID NO: 3
227 <211> LENGTH: 1775
228 <212> TYPE: DNA
229 <213> ORGANISM: Unknown
231 <220> FEATURE:
232 <223> OTHER INFORMATION: environmental source
234 <220> FEATURE:
235 <221> NAME/KEY: CDS
236 <222> LOCATION: (214)...(1491)
238 <400> SEQUENCE: 3
239 ggcaatttgc ggtgtttttt ccgcaggacg ttcacgtgcc ggctgtatt catcaacggc 60
240 cctgcgctat tcgcaaagtg gtggtgaaaa taccgctgcg ttatttaacg cccaataagc 120
241 aacaccgagt ttataaccct gaacgacacg gctgcggggc tgtgtagacg cccctacgcc 180
242 ttaacaccac taaatgactc tacaggtgta tat atg aat aca gcc tct gtt tct 234
243 Met Asn Thr Ala Ser Val Ser
244 1 5

```

RAW SEQUENCE LISTING

DATE: 10/22/2004

PATENT APPLICATION: US/10/787,267A

TIME: 12:50:01

Input Set : A:\GC687-3-D1-seqlist-corr.txt

Output Set: N:\CRF4\10222004\J787267A.raw

246	gtc acc caa agc cag gcg atc ccc aaa tta cgc tgg ttg aga ata gtg	282
247	Val Thr Gln Ser Gln Ala Ile Pro Lys Leu Arg Trp Leu Arg Ile Val	
248	10 15 20	
250	ccg cct att ctt att acc tgc att att tcc tat atg gac cgg gtg aac	330
251	Pro Pro Ile Leu Ile Thr Cys Ile Ile Ser Tyr Met Asp Arg Val Asn	
252	25 30 35	
254	atc gcc ttc gcc atg ccc ggc ggc atg gac gat gaa ctg ggc atc acc	378
255	Ile Ala Phe Ala Met Pro Gly Gly Met Asp Asp Glu Leu Gly Ile Thr	
256	40 45 50 55	
258	gcc tcg atg gcc ggg ttg gcc ggc ggt att ttc ttt atc ggt tat ctg	426
259	Ala Ser Met Ala Gly Leu Ala Gly Gly Ile Phe Phe Ile Gly Tyr Leu	
260	60 65 70	
262	ttc ttg cag gta ccc ggc ggc aag ctg gcg gtg tac ggc aac ggc aag	474
263	Phe Leu Gln Val Pro Gly Gly Lys Leu Ala Val Tyr Gly Asn Gly Lys	
264	75 80 85	
266	aaa ttc atc ggt tgg tcg ttg ttg gcc tgg gcg gtg att tcc gtg ctg	522
267	Lys Phe Ile Gly Trp Ser Leu Leu Ala Trp Ala Val Ile Ser Val Leu	
268	90 95 100	
270	acc ggg ctg gtc acg aat cag tat caa ttg ctg ttc ctg cgc ttc gcc	570
271	Thr Gly Leu Val Thr Asn Gln Tyr Gln Leu Leu Phe Leu Arg Phe Ala	
272	105 110 115	
274	ctc gcc cgt ttc cga agc ggc atg ctg cgg tgg gtg ctg acc atg atc	618
275	Leu Gly Arg Phe Arg Ser Gly Met Leu Arg Trp Val Leu Thr Met Ile	
276	120 125 130 135	
278	agc aac tgg ttc ccg gac aag gaa cgc ggg cgc gcc aac gcc atc gtc	666
279	Ser Asn Trp Phe Pro Asp Lys Glu Arg Gly Arg Ala Asn Ala Ile Val	
280	140 145 150	
282	atc atg ttc gtg ccg atc gcc ggc atc ctt acc gca ccg ctg tcc ggc	714
283	Ile Met Phe Val Pro Ile Ala Gly Ile Leu Thr Ala Pro Leu Ser Gly	
284	155 160 165	
286	tgg atc atc acc gcc tgg gac tgg cgc atg ctg ttc ctg gtc gag ggc	762
287	Trp Ile Ile Thr Ala Trp Asp Trp Arg Met Leu Phe Leu Val Glu Gly	
288	170 175 180	
290	gcg ctg tcg ctg gtc gtg atg gtg ctg tgg tat ttc acc atc agc aac	810
291	Ala Leu Ser Leu Val Val Met Val Leu Trp Tyr Phe Thr Ile Ser Asn	
292	185 190 195	
294	cgt cca caa gag gcc aaa agg att tcg cag gcg gaa aaa gat tat ctg	858
295	Arg Pro Gln Glu Ala Lys Arg Ile Ser Gln Ala Glu Lys Asp Tyr Leu	
296	200 205 210 215	
298	atc aaa acg ctg cac gac gaa cag ttg ctg atc aaa ggc aaa acg gtg	906
299	Ile Lys Thr Leu His Asp Glu Gln Leu Ile Lys Gly Lys Thr Val	
300	220 225 230	
302	cgc aac gcc tcg ctg cgt cgg gtg ctg ggc gac aaa atc atg tgg aag	954
303	Arg Asn Ala Ser Leu Arg Arg Val Leu Gly Asp Lys Ile Met Trp Lys	
304	235 240 245	
306	ttc ttc tac cag acc ggg ata tac ggc tac acc ctg tgg ctg ccg acc	1002
307	Phe Phe Tyr Gln Thr Gly Ile Tyr Gly Tyr Thr Leu Trp Leu Pro Thr	
308	250 255 260	
310	att ctc aag ggg ctc acc aac ggc aat atg gag cag gtc ggg atg ctg	1050

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/787,267A

DATE: 10/22/2004

TIME: 12:50:02

Input Set : A:\GC687-3-D1-seqlist-corr.txt

Output Set: N:\CRF4\10222004\J787267A.raw

L:1202 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:13
L:1205 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:13